

# INAUT, a Controlled Language for the French Coast Pilot Books *Instructions nautiques*\*

Yannis Haralambous, Julie Sauvage-Vincent, and John Puentes

Institut Mines-Télécom, Télécom Bretagne & UMR CNRS 6285 Lab-STICC  
Technopôle Brest Iroise CS 83818, 29238 Brest Cedex 3, France

**Abstract.** We describe INAUT, a controlled natural language dedicated to collaborative update of a knowledge base on maritime navigation and to automatic generation of coast pilot books (*Instructions nautiques*) of the French National Hydrographic and Oceanographic Service SHOM. INAUT is based on French language and abundantly uses georeferenced entities. After describing the structure of the overall system, giving details on the language and on its generation, and discussing the three major applications of INAUT (document production, interaction with ENC's and collaborative updates of the knowledge base), we conclude with future extensions and open problems.

## Introduction

*Instructions nautiques* is the name of a nautical book series [5] published by the French Marine Hydrographic and Oceanographic Service (SHOM). They are the French counterpart of the *United States Coast Pilot* [3], published by the United States National Oceanic and Atmospheric Administration's Office of Coast Survey, and of the British Admiralty Sailing Directions [7] published by the United Kingdom Hydrographic Office.

These publications aim to supplement charts (both paper ones and ENC's = Electronic Nautical Charts), in the sense that they provide the mariner with supplemental information not in the chart.

Information for the *Instructions nautiques* is provided by survey vessels, port officers, maritime officers and mariners in general. In some cases, it may require immediate update, for example to notify a shipwreck or some important change of the navigation conditions.

The SHOM is building a knowledge base that will cover both ENC's and nautical instructions. This knowledge base will communicate with ENC's and navigation equipment and, since updates can be frequent, the *Instructions nautiques* will have to be generated on-the-fly by the knowledge base.

To summarize, we have two constraints:

1. the information contained in the knowledge base, has to be easily updatable by people not necessarily proficient in the ontology formalism;

---

\* The final publication is available at <http://link.springer.com>.

2. the *Instructions nautiques*, or at least part of them, have to be automatically generated out of the knowledge base.

To fulfill constraint 1, we have built INAUT, a controlled language based on French natural language, and dedicated to the population and update of the SHOM knowledge base. Constraint 2 is fulfilled by generation of texts in INAUT out of the knowledge base. In fact, the texts generated will be in a more “literary” and concise version of the language, called LitINAUT (= Literary INAUT, §4) that will bring them closer to legacy human author production.

To our knowledge, INAUT is the first maritime CNL<sup>1</sup>

In the following, we present our model of the *Instructions nautiques* (§1), the SHOM knowledge base (§3), the controlled language (§3) and its generation (§4), the main operations (interaction with ENC’s (§5) and collaborative updates of the knowledge base (§6)) as well as future extensions and open problems (§7).

## 1 Modelling the *Instructions nautiques*

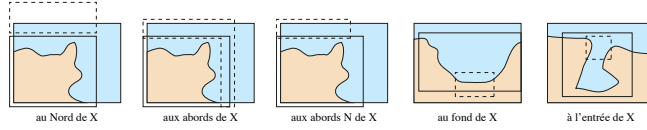
We model the *Instructions nautiques* as a set of three graphs (**S**, **G**, **K**): the *hierarchical structure of the document* **S**, the *geographic areas graph* **G** and the *SHOM knowledge base* **K** (see §3). Between these graphs we have two functions:  $g$  which maps some nodes of **S** and of **K** (those that are goereferenced) to nodes of **G**, and  $\kappa$  that maps leaf nodes of **S** to subgraphs of **K**. Furthermore, we have a set **T** of *titles of hierarchical subdivisions*, a set **A** of *geopolygons*, and functions  $\tau$  and  $\alpha$  mapping nodes of **S** (resp. **G**) to **T** (resp. **A**). Finally, there is a set **M** of functions  $\{\mu\}$  defined both as  $\mathbf{T} \rightarrow \mathbf{T}$  and as  $\mathbf{A} \rightarrow \mathbf{A}$ , called *modifiers*. Here are the details:

- graph **S** represents the hierarchical structure of a given volume. **S** is rooted, oriented and ordered. Let  $V(\mathbf{S})$  be the vertices and  $E(\mathbf{S})$  the edges of **S**;
- the five first levels of  $V(\mathbf{S})$  represent hierarchical subdivisions. Let us denote  $\ell$  the level function. The root  $n_0$  represents the entire document;
- function  $\tau: V(\mathbf{S}) \rightarrow \mathbf{T}$  maps every node  $n$  to a title  $\tau(n)$ ,
- $V(\mathbf{S})$  can be written as<sup>2</sup>  $V(\mathbf{S}) = V_G \sqcup V_{-G}$  ( $V_G$  are the *georeferenced nodes*) where we have a function  $g: V_G \rightarrow \mathbf{G}$  that maps every georeferenced node to a node in **G**, which again is mapped to a geopolygon in **A** by  $\alpha$ ;
- when generating a volume of the *Instruction nautiques*, the leafs of **S** are mapped to subgraphs of **K** through function  $\kappa$ . These subgraphs are then converted to text paragraphs in LitINAUT language;
- edges  $E(\mathbf{G})$  of **G** represent *partial inclusion*  $\dot{\subset}$  in **A**, in the sense that we have  $G'G \in E(\mathbf{G})$  (or  $a(G') \dot{\subset} a(G)$ ) if and only if  $\text{Area}(a(G') \cap a(G)) > 0.8 \text{Area}(a(G'))$ ;

<sup>1</sup> With the exception of Seaspeak [6], a CNL defined in 1985 by the International Maritime Lecturers Association. In 2001 it evolved into SMCP (Standard Marine Communication Phrases [2]) which is still used today. These CNLs, dedicated to oral communication between ships, are “human-only”.

<sup>2</sup> We denote by  $\sqcup$  the disjoint union:  $C = A \sqcup B \iff (C = A \cup B) \wedge (A \cap B = \emptyset)$ .

- the barycenters of  $\text{Im}\alpha \circ g$  when restricted to  $V_G \cup \ell^{-1}(i)$  for  $i \in \{1, 2, 3\}$  follow a path on the map, which corresponds to an itinerary along the coasts of France. The extremities of this itinerary for a given volume are given in  $t(n_0)$  where  $n_0$  is the root of  $\mathbf{S}$ . We call this path, the *guiding path* of the volume.
- modifiers  $\mu$  serve to describe locations relatively to other locations. For example, the modifier “au nord de X” (= to the North of X), applied to location “[cap Cerbère]” will produce “au nord de [cap Cerbère]” which is a new geographic entity, the polygon of which is calculated automatically out of  $\alpha(g([\text{cap Cerbère}]))$ . Some modifiers are shown in Fig. 1.



**Fig. 1.** The main modifiers: solid polygon represents the original area  $A$ , dashed polygon the modified one  $\mu(A)$ .

On Fig. 2 the reader can see an example of *Instructions nautiques* data in our model: on the left, the document structure tree, on the right, the geographical area graph. The gray box contains the LitINAUT text generated from section *Généralités* of §2.2.4. In the text, geographical entities are marked up by brackets. Dashed arrows between the two graphs represent function  $g$ .

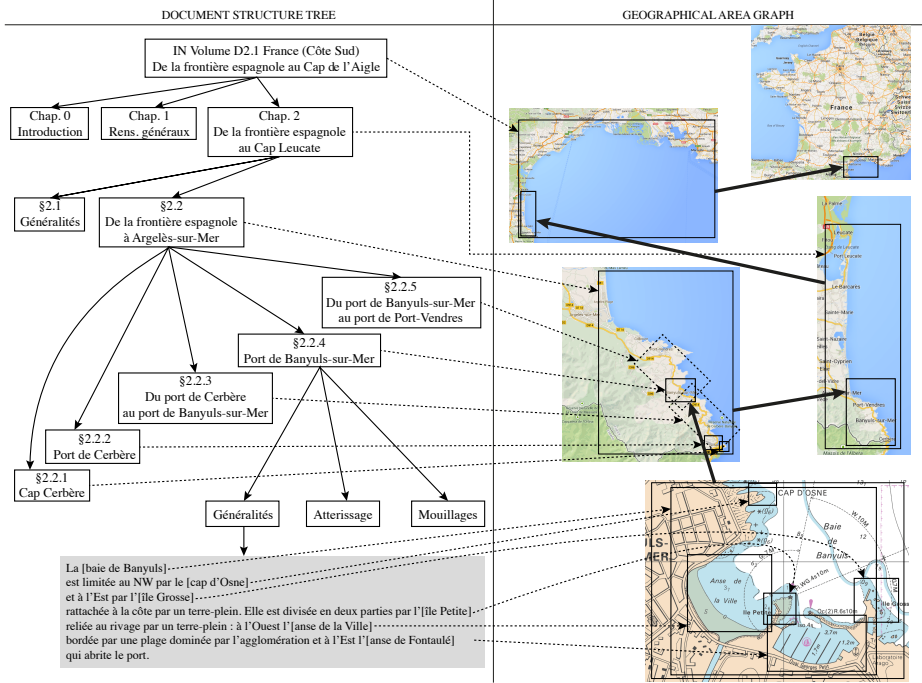
## 2 The SHOM knowledge base

Let us define (extending Cimiano [1]) a knowledge base  $\mathbf{K}$  as being a 16-tuple

$$(\mathfrak{C}, \leq_C, \mathfrak{A}, \mathfrak{R}_S, \mathfrak{R}_C, \mathfrak{T}, \mathfrak{I}, \mathfrak{V}, L_{\mathfrak{C}}, L_{\mathfrak{A}}, L_{\mathfrak{R}_S}, L_{\mathfrak{R}_C}, L_{\mathfrak{I}}, \sigma, \iota, \lambda)$$

where  $\mathfrak{C}, \mathfrak{A}, \mathfrak{R}_S, \mathfrak{R}_C, \mathfrak{T}, \mathfrak{I}, \mathfrak{V}$  are sets of concepts, attributes, simple relations, complex relations, types, instances and values,  $\leq_C$  is a hierarchy of concepts,  $L_{\mathfrak{C}}, L_{\mathfrak{A}}, L_{\mathfrak{R}_S}, L_{\mathfrak{R}_C}, L_{\mathfrak{I}}$  are sets of names of concepts, attributes, simple relations, relations and instances, and  $\sigma, \iota, \lambda$  denote signature, instantiation, lexicalization, as follows:

1.  $\iota: \mathfrak{C} \rightarrow 2^{\mathfrak{I}}$ ;
2. the signature of an attribute is  $\sigma: \mathfrak{A} \rightarrow \mathfrak{C} \times \mathfrak{T}$  and its instances  $\iota: \mathfrak{A} \rightarrow 2^{\mathfrak{I} \times \mathfrak{V}}$ ;
3. simple relations are relations between exactly two instances, without relation attributes. Hence we have  $\sigma: \mathfrak{R}_S \rightarrow \mathfrak{C} \times \mathfrak{C}$  and  $\iota: \mathfrak{R}_S \rightarrow 2^{\mathfrak{I} \times \mathfrak{I}}$ ;
4. complex relations are relations between  $n$  instances ( $n \geq 2$ ) which can also have relation attributes. Hence we have:  $\sigma: \mathfrak{R}_C \rightarrow \times^n \mathfrak{C} \times \times^m \mathfrak{T}$  (where  $\times^n$  denotes  $n$ -fold product) and  $\iota: \mathbf{R} \rightarrow 2^{\times^n \mathfrak{I} \times \times^m \mathfrak{V}}$ , with  $n \geq 2, m \geq 0$ ;



**Fig. 2.** Document structure tree and geographic area graph for an example taken from Vol. D2.1 of the *Instructions nautiques*.

5. an noteworthy difference between complex relations and simple relations, is lexicalization. Indeed, we have:  $\lambda: \mathfrak{C} \rightarrow L_{\mathfrak{C}}$ ,  $\lambda: \mathfrak{J} \rightarrow L_{\mathfrak{J}}$ ,  $\lambda: \mathfrak{A} \rightarrow L_{\mathfrak{A}}$ ,  $\lambda: \mathfrak{R}_S \rightarrow L_{\mathfrak{R}_S}$  as expected, but  $\lambda: \mathfrak{R}_C \rightarrow L_{\mathfrak{R}_C} \times \times^n L_{\mathfrak{R}_S} \times \times^m L_{\mathfrak{A}}$ , i.e., a relation has its own name, but requires also names for all instances involved in the relation as well as all relation attributes.

The concepts  $\mathfrak{C}$  of the SHOM knowledge base  $\mathbf{K}$ , belong to the domain of maritime navigation: ports, capes, sea currents, ships, etc.

As for  $\mathbf{S}$  nodes, instances  $\mathfrak{J}$  are of two types  $\mathfrak{J} = \mathfrak{J}_G \sqcup \mathfrak{J}_{-G}$ :  $\mathfrak{J}_G$  are georeferenced entities: “[baie de Banyuls]”, “[cap d’Osne]”, etc., in the sense that there is a map  $g$  between  $\mathbf{I}$  and the graph  $\mathbf{G}$ ;  $\mathfrak{J}_{-G}$  are non-georeferenced instances, such as “agglomération”, “port,” etc. They don’t need to be located on the map, and their purpose is purely descriptive of the environment.

Notice that the names of  $\mathfrak{J}_G$  instances often contain a hint to the predominant concept to which they belong (“baie”, “cap”, “port”, etc.), while in the case of  $\mathfrak{J}_{-G}$  instances, their names are often names of predominant concepts *per se*.

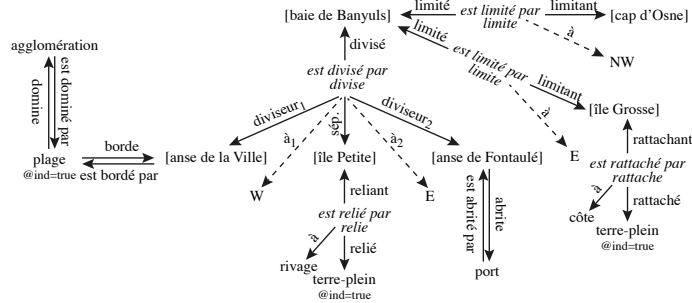
Simple relations  $\mathfrak{R}_S$  represent verbs in passive or active voice “est abrité par”, “est possible”, etc. Notice that most relations representing a passive verb have a symmetric relation representing the corresponding active verb: “A est abrité par B” has the symmetric relation “B abrite A”;

#### LitINAUT:

La [baie de Banyuls] est limitée au NW par le [cap d'Osne] et à l'Est par l'[île Grosse] rattachée à la côte par un terre-plein. Elle est divisée en deux parties par l'[île Petite] reliée au rivage par un terre-plein : à l'Ouest l'[anse de la Ville] bordée par une plage dominée par l'agglomération et à l'Est l'[anse de Fontaulé] qui abrite le port.

#### INAUT:

La [baie de Banyuls] est limitée au NW par le [cap d'Osne]. La [baie de Banyuls] est limitée à l'Est par l'[île Grosse]. L'[île Grosse] est rattachée à la côte par un terre-plein. La [baie de Banyuls] est divisée en deux parties par l'[île Petite] : à l'Ouest l'[anse de la Ville] et à l'Est l'[anse de Fontaulé]. L'[île Petite] est reliée au rivage par un terre-plein. L'[anse de la Ville] est bordée par une plage. La plage est dominée par l'agglomération. L'[anse de Fontaulé] abrite le port.



**Fig. 3.** A paragraph in INAUT, LitINAUT, and represented in the knowledge base.

Complex relations are  $n$ -ary ( $n \geq 2$ ) and can have attributes: for example “est limité par” has attribute “direction.” In this case, lexicalization requires names for all instances or attributes participating in the relation. In the case of “est limité par” the members of the relation are instances “limitant”, “limité” and attribute “à.”

As an example, the reader can see on Fig. 3 the sentence of Fig. 2, represented in LitINAUT, INAUT and as a subgraph of **K**. Instances “[baie de Banyuls]”, “[cap d’Osne]”, etc. belong to  $\mathcal{I}_G$ . Instances “côte”, “plage”, “port”, “rivage”, belong to  $\mathcal{I}_{-G}$ . Complex relations are reified as nodes. Attributes of instances have been included underneath, marked by character @.

### 3 The controlled languages INAUT and LitINAUT

INAUT is a controlled language with a rather large vocabulary (based on the existing *Instructions nautiques* corpus) but with a simple syntax, given by the following grammar:

$$\begin{aligned}
 S &\rightarrow NP \ VP \\
 NP &\rightarrow \text{MODIF} \ \text{DET} \ \text{NN} \mid \text{DET} \ \text{NN} \mid \text{NN} \\
 NN &\rightarrow \text{ADJ} \ \text{NN} \mid \text{NN} \ \text{ADJ} \ \text{NOUN} \\
 VP &\rightarrow \text{VERB} \ \text{NP} \mid \text{VERB} \ \text{NP} \ \text{PP} \\
 PP &\rightarrow \text{PREP} \ \text{DET} \ \text{NN} \mid \text{PREP} \ \text{NN}
 \end{aligned}$$

where symbols in small caps are terminal, all NOUNS belong to  $L_{\mathcal{I}}$  (the set of lexical references for instances of **K**) and to  $\mathcal{V}$  (the set of values of attributes of **K**), all VERBS belong to  $L_{\mathcal{R}_S} \cup L_{\mathcal{R}_C}$ , all ADJECTIVES belong to  $\mathcal{V}$ , and MODIFIERS, DETERMINANTS and PREPOSITIONS belong to a closed list.

The VERB, always in 3rd person or in the infinitive, can be active or passive. In most cases it is possible to change the voice of the verb, which implies a permutation of the NPs in SUBJECT and OBJECT position, leaving the PPs intact:

La [baie de Banyuls] est limitée par le [cap d’Osne] au NW.

Le [cap d’Osne] limite la [baie de Banyuls] au NW.

Definite articles are used for all instances in **G** the names of which start with the name of a concept to which belongs the instance: for example, the name “baie de Banyuls” starts with “baie” (= bay) which is the name of a concept in **C**, hence in INAUT the definite article is used: “la [baie de Banyuls]”.

Otherwise, no article is used:

[Notre-Dame de la Salette] est un amer remarquable à l’WSW du port.

Instances in  $\mathcal{J}_{-G}$  are, by default, used with definite articles. When an indefinite article is required, the information is stored in a dedicated attribute. Indefinite articles are used in OBJECT position only:

L’[anse de la Ville] est bordée par une plage. La plage est dominée par l’agglomération.

Modifiers are represented by (a closed set of) expressions outside the brackets of the geographic entity: in “au fond de l’[anse de la Ville]”, we have modifier “au fond de” and entity “[anse de la Ville].” In **K** there is a *modifier relation* whenever a modifier is used. This relation does not produce INAUT text but serves to connect subgraphs in **K** during content determination (§4).

In the following sections we will discuss the three main operations of controlled languages INAUT and LitINAUT: generation (§4), interaction with ENC’s (§5), collaborative updates of the knowledge base (§6).

## 4 Controlled language generation

One of the design goals of our system is to be able to produce automatically a large part of the *Instructions nautiques*, so that after collaborative updates new versions of the entire document can automatically be produced.

We have divided the task into two stages: (1) produce INAUT text corresponding to a given leaf node of **S**; (2) convert INAUT language into LitINAUT.

Suppose given a leaf node *S* of **S**. Producing the corresponding INAUT text is typically a Natural Language Generation problem.

Reiter & Dale [4, §3.3] divide the language generation task into seven sub-tasks: content determination, document structuring, lexicalisation, aggregation, referring expression generation, linguistic realisation and structure realisation.

*Content determination.* Using Algorithm 1, we find the subgraph *K* of **K** which is geographically the most relevant to *S*. We apply tags to its connected components using a rule-based decision system: for example, when a connected component contains the instance “mouillage” then it is tagged as belonging to a leaf node of type “Mouillages.” If after applying the rules no tag has been affected, then the component belongs to a leaf node of default type “Généralités.”

**Input:**  $\mathbf{S}, \mathbf{G}, \mathbf{K}$  and a leaf node  $S \in \mathbf{S}$   
**Result:** The subgraph  $K \subset \mathbf{K}$  which represents the text corresponding to  $S$   
 $G_S \leftarrow g(S)$ ;  
 $K \leftarrow \emptyset$ ;  
**for**  $G \in \mathbf{G}$  **do**  
    **if**  $G \dot{\subset} G_S$  and  $g^{-1}(G) \in \mathcal{I}_G$  **then**  
         $K \leftarrow K \cup g^{-1}(G)$ ;  
    **end**  
**end**  
**for**  $k \in K$  **do**  
    **if**  $\exists k'$  such that  $kk' \in \text{Undirected}(\mathbf{K})$  and  $kk' \notin K$  **then**  
        **if**  $k' \in \mathcal{I}_{-G}$  or  $k' \in \mathcal{V}$  or  $(k' \in \mathcal{I}_G$  and  $g(k') \dot{\subset} G_S)$  **then**  
             $K \leftarrow K \cup kk'$ ;  
        **end**  
        **if**  $k' \in \mathcal{R}_C$  and  $\exists k''$  member of  $k'$  such that  $g(k'') \dot{\subset} G_S$  **then**  
             $K \leftarrow K \cup kk'$ ;  
            **for**  $m$  member of  $k'$  **do**  
                 $K \leftarrow K \cup k'm$ ;  
            **end**  
        **end**  
    **end**  
**end**

**Algorithm 1:** Content determination algorithm.

*Document structuring.* This is the most difficult phase since it deals with the order in which sentences are written. Let  $K$  be the subgraph of  $\mathbf{K}$  to be converted into INAUT.

We subdivide the task in four subtasks:

1. sort connected components  $K_i$  of  $K$ ;
2. for each component find a starting node  $s$ ;
3. find the order in which the relations of each component will be converted into INAUT, starting from  $s$ ;
4. convert relations into INAUT in the order given by 1 and 3.

For subtask 1, we will sort components. The sorting criteria are: (a) if there is a significant difference in size between the cumulated geographic areas of two components, the larger one will precede the smaller one, (b) otherwise, calculate the barycenters of cumulated geographic areas of components; the path defined by their barycenters should be roughly parallel to the guiding path of the volume. For example, on Fig. 2 the areas of nodes §2.2.1–§2.2.5 follow a SE to NW direction, this direction can be chosen for the order of connected components.

To accomplish steps 2 and 3 we define weights  $w$  on nodes and relations. Calculation of these weights is based on criteria we will describe below, as well as on training using machine learning algorithms on the existing *Instructions nautiques* corpus.

Notice that we use undirected graphs since every edge can be inverted by changing voice.

The first and most obvious criterion is the relation between nodes in  $K_i$  and the parent of the leaf node of  $\mathbf{S}$  that established the connection with  $\mathbf{K}$  (i.e.,  $\kappa^{-1}(K)$ ). If among the nodes there is one whose geographic area and/or name

matches as closely as possible the one of the parent of the leaf node, it is a good choice. For example, in our case, node §2.2.4 of **S** is “Port de Banyuls-sur-Mer” which is much closer to “baie de Banyuls” than to “cap d’Osne”, both in terms of geographic area than simply of string comparison of names.

The second criterion for choosing the starting node is its position in the **G** lattice. Let  $k_m = g^{-1}(\max_{\mathbf{G}} g(K_i))$ . If  $k_m \in K_i$  then it is an obvious choice. Otherwise we take local maxima in **G** and proceed with weighting.

Finally, another criterion is of semantic nature, the one of “interest” for the navigator: an order can be established between concepts to which instances of  $K_i$  belong, for example a port instance will be more interesting than a beach instance. This weight is inherited by neighboring nodes: for instance, a bay containing a port is more interesting than a bay containing a beach, etc.

The “semantic weight” of instances can be calculated by machine learning.

Once the starting point has been established, we proceed to subtask 3. We will use a variant of DFS (depth-first search) to search  $K_i$ .

Subtask 4 is the simplest one: from relations in **K** we build INAUT sentences, by applying rules, for example choosing the verb’s voice according to the direction of the search in  $K_i$ , adding articles matching nouns, etc. There still remains a difficulty: finding the right order of prepositional phrases, as in:

La [baie de Banyuls] est limitée par le [cap d’Osne] au NW.

La [baie de Banyuls] est limitée au NW par le [cap d’Osne].

When the difference may be purely stylistic (as above), the order can be obtained by machine learning. In other cases, such as in “est divisé par” of Fig. 2, it is mandatory to group some relations: the text representing “à<sub>i</sub>” must immediately follow the one representing “diviseur<sub>i</sub>” since indices disappear in the textual realization and only proximity allows to distinguish the divisors of the entity.

Another important phenomenon is text added by default: for example, to realize relation “est divisé par” we need to add the number of divisors, this is done by counting the members of the relation of type “diviseur” and generate “en deux parties” (= in two parts).

*Aggregation and referring expressions generation: LitINAUT language.* At this stage, generation of INAUT has been completed. The result, as it can be seen in Fig. 3, is not very eloquent, but remains closely related to the structure of **K**, so that it is easier for contributors to supply modifications and additions written in INAUT. To produce a human readable text as part of automatically generated *Instructions nautiques* document, we need two extra steps: aggregation of several sentences into a single one, and generation or referring expressions. The result of these two operations is called LitINAUT language.

By the fact of using DFS to search  $K_i$ , often the object of a sentence is the subject of the following one. Aggregation merges them into a single sentence:

L[anse de la Ville] est bordée par une plage.

La plage est dominée par l’agglomération.

→ L[anse de la Ville] est bordée par une plage, dominée par l’agglomération.



In other cases, consecutive sentences have the same object and the same verb; in that case we use conjunction:

La [baie de Banyuls] est limitée au NW par le [cap d’Osne].  
La [baie de Banyuls] est limitée à l’Est par l’[île Grosse].  
→ La [baie de Banyuls] est limitée au NW par le [cap d’Osne] et à l’Est par l’[île Grosse].

When we have object identity but with different verbs, referring expressions are generated:

La [baie de Banyuls] est divisée en deux parties par... → Elle est divisée en deux parties par...

In some cases, text is omitted from the realization because it is obtained from the context: for example, in realizing the text corresponding to a subdivision of type “Mouillages” (= mooring), we will systematically omit the part “Le mouillage est autorisé à” since it is implied by the subdivision title.

These are just some examples of mechanisms used to convert INAUT into LitINAUT. Work is in progress to enhance the result and bring it closer to legacy (human authored) text.

## 5 Interaction with ENC’s

As said in the introduction, *Instructions nautiques* are defined as a complement to charts, and, in particular, to ENC’s. Therefore it is important to define interactions between **K** and ENC’s, via INAUT. By specifying, for example, an area of interest on an ENC (for example, by drawing a zone on a touchscreen) the user may receive LitINAUT text in return. Generating this text automatically has the advantage of being (a) limited to the zone of interest given by the user; (b) conform to local conditions, for example time of the day (some relations or attributes in **K** may be time-dependent) or meteorological conditions, or parameters of the user’s vessel (size, tonnage, etc.); (c) up-to-date, since other users may constantly provide new information.

To provide adaptive LitINAUT text, we first position the area  $U$  given by the user in **A**, and hence in **G**. Knowing the subgraph of **G** that matches as closely as possible  $U$  in **A**, we find the relevant nodes in **K** by going through  $g^{-1}$ . These nodes form a subgraph of **K** and we apply the techniques described in 4 to generate the corresponding text in LitINAUT.

Additional structure can be added to the text sent to the ENC device, so that the user can filter the text and display only specific types of information, as for example information on mooring, landing, etc.

## 6 Collaborative updates of the knowledge base

It is important for the SHOM knowledge base to be kept constantly up-to-date. To achieve this goal, INAUT will be used as a tool for collaborative update.

Indeed, INAUT has been designed as the optimal compromise between easiness of use (since contributors have a priori no KM proficiency) and formality (as the knowledge base will be fed directly by the incoming data).

To make the system more robust, we validate on two levels. First, the lexical and syntactic level: a Web interface analyzes segments written in INAUT and validates them. In case of errors it provides correction hints. Second, the semantic level: a human controller monitors incoming INAUT data which, depending on the contributors trust level are automatically fed into the knowledge base (with the possibility of making the modification retroactive) or are stored in a waiting list until manual validation.

## 7 Conclusion and future work

We have described the controlled natural language INAUT (and its variant LitINAUT) which is used for the update of the SHOM maritime knowledge base, for automatic generation of *Instructions nautiques* documents and for interaction with ENC's.

Among our plans is the extension of INAUT into a QA system. This requires extension of INAUT to interrogative sentences and increased use of the concept hierarchy. Another extension deals with the issue of *dangerosity*. Indeed, one of goals of *Instructions nautiques* is to alert the navigator on possible dangers. Ideally, the ENC should automatically send queries about dangerosity to the knowledge base involving the current position of the vessel and various external conditions, and in case of a positive answer, alert the navigator by all means possible. Special NLG techniques can then be used, since the communicative goal will not be simply to inform, but to alert.

## References

1. Cimiano, P.: *Ontology Learning and Population from Text. Algorithms, Evaluation and Applications*. Springer (2008)
2. IMO: *IMO Standard Marine Communication Phrases, with pronunciation*. International Maritime Organization (2005)
3. NOAA: *United States Coast pilot*, <http://www.nauticalcharts.noaa.gov/nsd/cpdownload.htm>
4. Reiter, E., Dale, R.: *Building Natural Language Generation Systems*. Cambridge University Press (2000)
5. SHOM: *Instructions nautiques*, <http://www.shom.fr/les-produits/produits-nautiques/ouvrages-nautiques/instructions-nautiques/>
6. Strevens, P., Johnson, E.: SEASPEAK: a project in applied linguistics, language engineering, and eventually ESP for sailors. *The ESP Journal* 2(2), 123–129 (1983), <http://www.sciencedirect.com/science/article/pii/0272238093900020>
7. UKHO: *Admiralty sailing directions*, <https://www.ukho.gov.uk/PRODUCTSANDSERVICES/PAPERPUBLICATIONS/Pages/NauticalPubs.aspx>